Effective Feature Preprocessing for Time Series Forecasting

Jun Hua Zhao[1], ZhaoYang Dong[1], and Zhao Xu[2]

[1] The School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, QLD 4072, Australia
{zhao, zdong}@itee.uq.edu.au
[2] Centre for Electric Technology (CET), Ørsted*DTU, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
zx@oersted.dtu.dk

Abstract. Time series forecasting is an important area in data mining research. Feature preprocessing techniques have significant influence on forecasting accuracy, therefore are essential in a forecasting model. Although several feature preprocessing techniques have been applied in time series forecasting, there is so far no systematic research to study and compare their performance. How to select effective techniques of feature preprocessing in a forecasting model remains a problem. In this paper, the authors conduct a comprehensive study of existing feature preprocessing techniques to evaluate their empirical performance in time series forecasting. It is demonstrated in our experiment that, effective feature preprocessing can significantly enhance forecasting accuracy. This research can be a useful guidance for researchers on effectively selecting feature preprocessing techniques and integrating them with time series forecasting models.

1 Introduction

Time series forecasting is a major challenge in many real-world applications, such as stock price analysis [1], electricity price forecasting [2] and flood forecasting [3]. Time series forecasting is to predict the values of a continuous variable (called *response variable*) with a forecasting model based on historical data. Given its valuable potential in many applications, time series forecasting has been a hot research topic since the past decade. There are many time series models developed for this purpose, including ARIMA [2] and Fuzzy-neural autoregressive models [4]. Neural Networks can also be used to approximate time series data [5]. Garch model [6] is proven to have good performance especially in handling the volatility in some time series such as stock prices. Recently, Support Vector Machine (SVM) has also been applied in time series forecasting and achieved satisfactory results [7].

Given a specific forecasting model, feature preprocessing is essential for improving forecasting accuracy. *Feature preprocessing* [8] is the process of selecting the relevant factors of a response variable, and/or constructing new features based on these factors. The following two reasons make feature preprocessing very important for time series forecasting. Firstly, training the forecasting model with irrelevant features can greatly degrade the forecasting accuracy, and also decrease the training

X. Li, O.R. Zaiane, and Z. Li (Eds.): ADMA 2006, LNAI 4093, pp. 769 – 781, 2006.

speed to an intractable level; moreover, when there are too many available features, manually determine the relevant features of the response variable with statistical plots will become impossible. Therefore, effective methods are required to automatically choose the most relevant feature set. This process is known as *feature selection* [9]. Secondly, the correlations between different features can seriously affect the performance of the forecasting model. *Feature extraction* [10] techniques can be employed to construct the new features that are mutually independent, and reduce the noise in data.

Feature selection/extraction have been extensively studied in statistics [11], machine learning [12] and data mining [13], and widely applied to many areas. In [14-17], several feature selection techniques are proposed as a preprocessing tool selecting the relevant feature subset. Feature extraction techniques can be applied to generate independent features which still maintain the relationships between original features and response variable. These techniques include principal component analysis (PCA) [18] and independent component analysis (ICA) [19]. Signal processing techniques such as wavelet decomposition and reconstruction [20], have also been applied to reduce the noise in training data. Some of these techniques have been applied in time series forecasting problem [20-22]. However, no systematic research has been performed yet to study the general performance of different feature preprocessing techniques. Without empirical studies, it is difficult to determine the most effective and proper feature preprocessing techniques for time series forecasting.

In this paper, the authors attempt to conduct a comprehensive study of existing feature preprocessing techniques, and evaluate their empirical performance in time series forecasting. We will firstly give a brief introduction to the general procedure of feature preprocessing and several existing feature selection/extraction techniques. A comprehensive empirical study will be given to compare the performance of different techniques on real-world datasets. The main contribution of this paper is to present a comprehensive study on how to choose the most suitable feature preprocessing techniques based on the characteristics of data and the requirements of time series analysis.

The rest of the paper is organized as follows. The definitions of time series forecasting and feature selection/extraction are presented in Section 2. A systematic introduction to feature preprocessing, including the general procedure and several widely used techniques are given subsequently. In Section 4, Support Vector Machine is briefly reviewed as the forecasting model for our case studies. In Section 5, a comprehensive empirical study of feature preprocessing is given. Finally, Section 6 concludes the paper.

2 Problem Formulation

Before we discuss time series forecasting and its feature preprocessing, clear definitions of forecasting and feature selection/extraction should be firstly given. Time series forecasting techniques are studied in several different areas, including statistics, machine learning, and data mining. Generally, forecasting is considered as a *supervised learning* problem [23], and can be solved by regression [23] or time series techniques [23]. The formal definition of forecasting is given as follows:

Definition 1: Given a historical time series dataset $S = \{(X_1, y_1), (X_2, y_2)...(X_t, y_t)\}$, where y_t is the *response variable* to be forecast at time *t*, and $X_t = (x_{t1}, x_{t2}, ...x_{tm})$ represents the values of relevant factors. We assume that a function dependency $f : S \rightarrow y_{t+n}$ exists, where y_{t+n} denotes the response variable *n* time units later than time *t*. Then the *forecasting* problem is to find a proper f' approximating f, and use f' to forecast the values of the response variable in the future.

Because training the forecasting model with the original feature set may not reach the optimal forecast accuracy, feature selection and extraction techniques are proposed to obtain a better feature set. Their definitions are given as follows:

Definition 2: Given a historical time series dataset $S = \{(X_1, y_1), (X_2, y_2)...(X_t, y_t)\}$, and denote $X = \{x_1, x_2, ...x_m\}$ as the original feature set. *Feature selection* is the process of generating an optimal feature subset $X' \subseteq X = \{x_1, x_2, ...x_m\}$, based on which f' will have the optimal forecast accuracy on the future dataset.

A simple illustration of feature selection is given in Fig.1. Clearly, among the many features available, only a small number of relevant features are identified through feature selection.

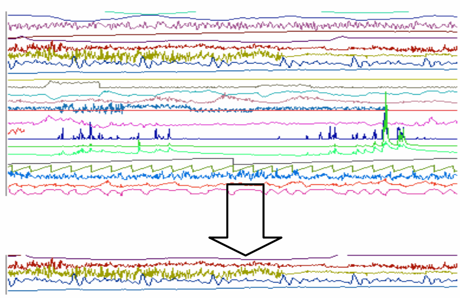

Fig. 1. An example of feature selection

Definition 3: Given the historical time series dataset $S = \{(X_1, y_1), (X_2, y_2)...(X_t, y_t)\}$, and denote $X = \{x_1, x_2, ...x_m\}$ as the original feature set. *Feature extraction* is a process to generate the optimal feature set $X' = (x_1'(x_1, ...x_m), x_2'(x_1, ...x_m)...x_l'(x_1, ...x_m))$, based on which f' will have the optimal forecast accuracy on the future dataset. Here each new feature in X' is a function of the original features.

Obviously, feature selection, which selects a subset of the original feature set, is a special case of feature extraction. Feature extraction transforms the original feature set into another feature space. However, the optimal forecast accuracy is the objective of both feature selection and extraction.

3 Feature Preprocessing Techniques

In this section, the authors will first introduce the basic procedure of feature preprocessing before discussing some well-known feature preprocessing techniques. We will also discuss how to choose the appropriate feature preprocessing techniques to suit the specific requirements of time series data analysis.

3.1 Procedure of Feature Preprocessing

Feature preprocessing usually includes four steps as follows [9]:

- **Candidate set generation.** Candidate set generation is a search procedure that uses a certain search strategy to produce candidate feature sets. Here the candidate feature sets can be the subsets of the original feature set (feature selection), and also can be the transformations of the original feature set (feature extraction). Feature preprocessing techniques can be classified according to the different search strategies in this step.
- **Candidate set evaluation.** An evaluation criterion should be applied to determine whether each candidate feature set is better than the previous best one. If the new feature set is superior, it replaces the previous best feature set. According to the different evaluation models employed in the step of candidate set evaluation, feature preprocessing techniques can also be categorized as the *filter model*, *wrapper model* and *hybrid model* [9].
- **Stopping criterion.** The process of candidate set generation and evaluation is repeated until a given stopping criterion is satisfied.
- **Results validation.** The selected optimal feature set usually needs to be validated by prior knowledge or different tests with synthetic and/or real-world data sets.

The complete feature preprocessing procedure and the relationship between the four major steps are given in Fig. 2.

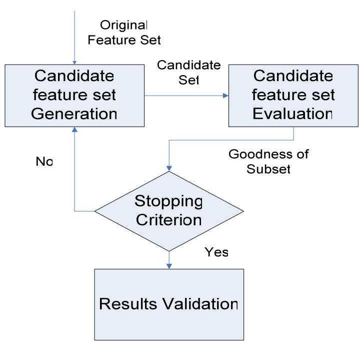

Fig. 2. Feature preprocessing procedure

3.2 Some Well-Known Feature Preprocessing Techniques

In this subsection, several existing feature preprocessing techniques will be discussed in details. These techniques employ different search strategies in candidate set generation, and different evaluation criteria in candidate set evaluation. We will discuss several combinations of search strategy and evaluation criterion respectively. Their effectiveness in time series forecasting will be demonstrated in experiments section.

The first step of feature preprocessing is to generate candidate feature sets. To guarantee identifying the optimal candidate feature set, a straightforward method is to perform an exhaustive/complete search on all possible feature subsets. However, this is computationally expensive especially when dealing with a large number of original features. Consequently, several heuristic search strategies are proposed as follows:

- **Complete Search [9]**. As mentioned above, complete search is usually intractable, although it guarantees the global optimum. This search strategy is seldom used in real-world applications.
- **Best-first Search.** Best-first strategy searches the space of feature subsets by *greedy hill climbing* [24], and employs a backtracking facility to avoid the local optima. It may start with the empty set and search forward, or with the full set of attributes and search backward, or start at any point and search in both directions. Because of its heuristic strategy, there is no guarantee of the global optimum.
- **Greedy-stepwise Search.** This search strategy performs greedy forward or backward search, starting from an arbitrary point in the feature space [25]. It can produce a ranked list of attributes by traversing the feature space from one side to the other and recording the order that attributes are selected. This strategy is also locally optimal.
- **Genetic Search.** Genetic search employs the simple genetic algorithm to locate the global optimum [26]. Theoretically this strategy has the capability of locating the global optimum. Meanwhile its search speed is not significantly slower than best-first and greedy-stepwise methods.
- **Random Search.** Random search generates each candidate set without any deterministic rule [9]. The users can predetermine the percentage of the feature space to be explored. Theoretically this strategy is globally optimal if a large percentage is set, which usually makes the search process intractable.
- **Ranker.** When a predetermined evaluation criterion can be employed to evaluate each attribute individually, we may simply rank the features according to their values given by the criterion. This strategy is the fastest search strategy. However, when the features are correlated, results of ranker strategy are usually unreliable.
- **Feature transformation.** Different from the above search strategies, feature transformation generates new feature sets according to the characteristics of data. For example, *principle component analysis (PCA)* [18] chooses enough eigenvectors to account for some percentages of the variance in original data (e.g. 95%). Attribute noise can be filtered by transforming original features to the principal component (PC) space. Other feature transformation algorithms include the *Fourier transformation* [27], *wavelet transformation* [20], *independent component analysis (ICA)* [19], etc.

After the candidate feature sets are generated, a number of evaluation criteria can be used to evaluate the quality of feature sets. These criteria include:

- **Distance measure [9].** Distance measures are also known as *separability*, *divergence*, or *discrimination* measures. For different values of a response variable, a feature X is considered more relevant than another feature Y if X produces a greater difference between the conditional probabilities than Y. A famous feature selection technique employing distance measure is *relief* [16].

- **Information measure [15].** Information measures calculate the *information gain* [15] of each feature. The information gain from a feature X is defined as the difference between the prior uncertainty and expected posterior uncertainty considering X. Features with large information gain are considered as good features. There are other information measures, such as *gain ratio* [28].
- **Dependency measure.** Also known as *correlation* measures or *similarity* measures, dependency measure evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Subsets of features are preferred if they are highly correlated with the response variable and have low inter-correlation. *CFS* [14] is a popular dependency measure.
- **Consistency measure [9].** Given different values of a response variable, consistency measures aims at searching a minimum number of features, which have a predictive ability similar to the full set of features. Two instances are inconsistent if they have the same feature values but different values of the response variable.

The above four measures belong to the *filter* model [9]. They evaluate the quality of feature sets according to predetermined measures. These measures are independent of a certain forecasting model. However, filter models may not lead to the highest forecast accuracy, because they do not guarantee to be consistent with accuracy criteria. To tackle this shortcoming, the following two models can be used:

- **Wrapper model [17].** Different from the filter model, the wrapper model integrates a specific forecasting model (e.g. SVM) with the feature preprocessing process. After each candidate feature set is generated, wrapper model trains the forecasting model with this feature set and obtains the forecasting accuracy. This accuracy will then be used as the quality criterion of feature sets in the second step of feature preprocessing. The wrapper model usually outperforms the filter model, because it guarantees to generate the feature set leading to highest forecasting accuracy. However, the wrapper model requires a training time much longer than the filter model.
- **Hybrid model [9].** Hybrid model employs different criteria of both filter and wrapper models at different search stages to overcome the limitations of filter model and wrapper model.

Given the feature preprocessing techniques discussed above, users may select different methods according to different purposes and requirements. First, several evaluation measures, such as the distance, information and consistency measures, are designed to handle discrete response variables. Therefore, discretization [23] should be performed on data before these measures can be applied in time series forecasting. Second, feature extraction methods should be applied considering the quality of data. For example, PCA can be applied when original features are highly correlated, while noisy data can be filtered with wavelet transformation.

4 Support Vector Machine

S*upport Vector Machine (SVM)* will be used in the experiments as the forecasting model. We give a brief introduction to SVM for completeness.

SVM is a new machine learning method developed by Vladimir Vapnik et al at Bell Laboratories [29]. This method received increasing attention in recent years because of its excellent performance in both classification and regression. It has been proven that SVM has excellent performance in time series forecasting problems [7].

The simplest form of SVM is the *linear regression,* which can be used for linear training data. For the training data $\{(X_1, y_1),......(X_l, y_l)\} \subset R^n \times R$, we can use Vapnik's ε-insensitive loss function

$$|y - f(X)|_\varepsilon := \max\{0, | y - f(X) | - \varepsilon\} \tag{1}$$

to estimate a linear regression function

$$f(X) = < W \bullet X > +b. \tag{2}$$

with precision ε. This linear regression problem can be solved by minimizing

$$\frac{\|W\|^2}{2} + C\sum_{i=1}^{m} |y_i - f(X_i)| \tag{3}$$

It is equivalent to a constrained optimization problem of (4)-(5):

$$\text{Minimize} \quad \frac{\|W\|^2}{2} + C\sum_{i=1}^{m} |\xi_i + \xi_i^*| \tag{4}$$

$$\text{Subject to} \quad \begin{aligned} &(< W \bullet X_i > +b) - y_i \le \varepsilon + \xi_i \\ &y_i - (< W \bullet X_i > +b) \le \varepsilon + \xi_i^* \\ &\xi_i, \xi_i^* \ge 0 \end{aligned} \tag{5}$$

The *Lagrange multipliers method* can be used to solve this optimization problem.

In most of real-world problems, the relationship between *y* and *X is* not linear. One method to deal with non-linear data is to use a map function $\Phi(X): R^n \mapsto H$ to map the training data from input space into some high dimensional feature space where the data become linear. SVM can then be applied in the feature space. Note that the training data used in SVM are only in the form of dot product, therefore, after the mapping the SVM algorithm will only depend on the dot product of $\Phi(X)$. If we can find a function that can be written in the form of $K(X_1, X_2) =< \Phi(X_1), \Phi(X_2)>$, the mapping function $\Phi(X)$ will not need to be explicitly calculated in the algorithm. $K(X_1, X_2)$ is a *kernel function* or *kernel*. Radial basis kernel [29] is used in this paper:

$$K(X,Y) = \exp(-\frac{\|X - Y\|^2}{2\sigma^2}) \tag{6}$$

5 Experiment

We conduct a comprehensive experiment to study the effectiveness of several feature preprocessing techniques in the time series forecasting problem. The electricity price dataset of the *National Electricity Market (NEM)* of Australia is used for the experiment. The market operator NEMMCO operates this competitive electricity market and publishes the historical and real-time data of NEM *regional reference price* (*RRP*) at its website. Electricity price forecasting is known as a challenging time series forecasting problem because of the high volatility of electricity prices [6]. The data of electricity prices are therefore used as the experiment data in this paper.

In the experiment, 15 feature preprocessing techniques will be firstly applied on the NEM price data of Jan, 2004, and generate 15 feature sets. These 15 feature sets will be employed to train 15 SVM regression models separately. Finally these 15 SVM models will be tested with the NEM price data of Jan, 2005. The forecast accuracy achieved by these 15 models represents the quality of the feature sets produced by corresponding feature preprocessing techniques.

Mean absolute percentage error (MAPE) is a commonly used criterion of forecast accuracy, due to its robustness and simplicity. MAPE is defined as:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\times 100\right| \tag{7}$$

In our experiment, MAPE will also be used as a measure of forecast accuracy and feature set quality.

The real price data of NEM, which originally include 48 features, are used in the experiment. We firstly train a SVM model with all 48 features. The data of Jan, 04 is chosen as the training data, while the data of Jan, 05 is the test data. The forecasting result is shown in Fig. 4.

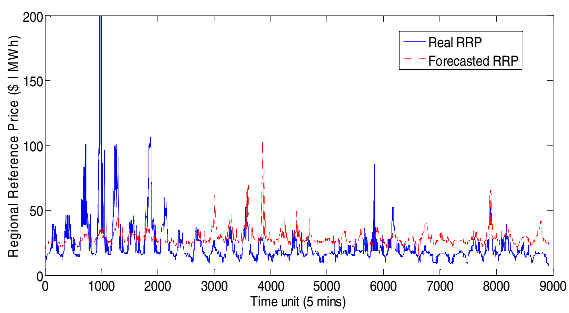

Fig. 4. Price forecast with original feature set

As shown in Fig.4, the forecast with the original feature set has large errors. The MAPE is 42.84%, which is not satisfactory.

In the second step of the experiment, 15 different feature preprocessing techniques are applied to generate 15 different feature sets. 15 SVM models are then trained separately with these 15 feature sets. Real market data of Jan, 04 and Jan, 05 are still used as the training and testing data. The accuracy achieved is listed as below:

Table 1. Forecast accuracy of 15 feature preprocessing techniques

Index	Search/Generation Strategy	Evaluation Criterion	Number of features generated	MAPE (%)
1	Best-First	CFS	2	24.51
2	Genetic Search	CFS	14	15.34
3	Random Search	CFS	11	25.24
4	Greedy Stepwise	CFS	2	24.51
5	Ranker	Chi-square measure [25]	12	14.04
6	Best-First	Consistency measure	2	24.51
7	Greedy Stepwise	Consistency measure	10	29.04
8	Genetic Search	Consistency measure	10	23.05
9	Ranker	Gain ratio	15	22.19
10	Ranker	Information Gain	13	20.27
11	Ranker	OneR Attribute Evaluation [25]	10	26.62
12	N.A.	PCA	13	13.17
13	Ranker	Symmetrical Uncertainty [25]	18	22.47
14	Ranker	Relief	15	12.78
15	Genetic Search	Wrapper Model	28	9.46

As shown in Table 1, every feature preprocessing technique leads to a smaller MAPE than that of the original feature set. This means that including irrelevant features in the forecasting model can significantly decrease the forecasting accuracy.

Among the 15 techniques, wrapper model + genetic search is the best approach with a MAPE < 10%. Moreover, relief + ranker, PCA, chi-square measure + ranker and CFS + genetic search also have good performance, which are within 15%.

The average accuracy of different search/generation strategies is listed in Table 2. Obviously, genetic search has a better performance than other alternatives. A surprising phenomenon is that, random search performs poorly, although it is theoretically global optimal. An explanation is that, to speed up the random search, we usually set a small percentage of the feature space to be explored. For example, in our experiment, we only search $1/10^6$ of the feature space. However, the random search costed more than 24 hours on a PC with Pentium IV CPU and 512M memory. Since only a small region of the feature space has been searched, random search usually cannot locate the global optimum.

Table 2. Average accuracy of different search/generation strategies

Search/Generation Strategy	Average number of features generated	Average MAPE (%)
Best-First	2	24.51
Genetic Search	17	15.95
Random Search	11	25.24
Greedy Stepwise	6	26.78
Ranker	14	19.73

The results of Table 2 are also visualized in Fig. 5. As is shown, methods that generate fewer features usually produce larger errors. This can be explained as, besides the irrelevant and noisy features, some relevant features are also removed by those feature preprocessing methods, which accounts for the performance degradation. Therefore, the optimal feature preprocessing technique should remove all irrelevant features while still keeping the features with useful information.

The average accuracy of different evaluation criteria is also listed in Table 3. The Wrapper model is the best evaluation criterion, although it is also the most time-consuming. Within the filter models, Chi-square measure and Relief have better performance. The MAPE of PCA is also at a low level, which means that severe feature correlations exist in the price data.

The results of Table 3 are illustrated in Fig. 6. Similarly, wrapper model, which has the largest feature number, has the smallest MAPE.

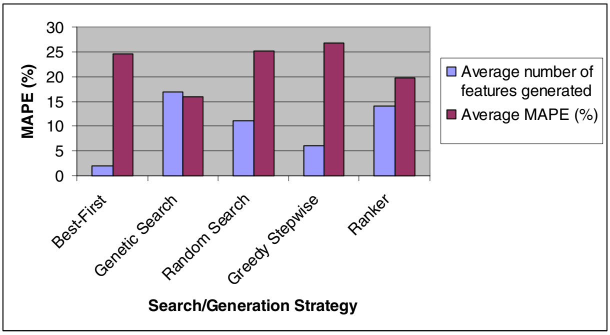

Fig. 5. MAPE and feature numbers of different search/generation strategy

To understand the importance of feature preprocessing, the price forecasts using wrapper model + genetic search, relief + ranker and PCA, are plotted in Figs. 7-9.

Table 3. Average accuracy of different evaluation criteria

Evaluation criterion	Average number of features generated	Average MAPE (%)
CFS	7	22.4
Chi-square measure	12	14.04
Consistency measure	7	25.53
Gain Ratio	15	22.19
Information Gain	13	20.27
OneR Evaluation	10	26.62
PCA	13	13.17
Symmetrical Uncertainty	18	22.47
Relief	15	12.78
Wrapper	28	9.46

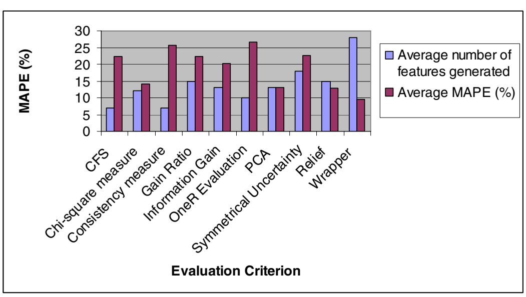

Fig. 6. MAPE and feature numbers of different evaluation criteria

Comparing Fig. 4 with Figs. 7-9, we can clearly observe that the performance of SVM is significantly improved after effective feature preprocessing techniques are integrated. This clearly demonstrates that feature preprocessing is an important part of time series forecasting. Moreover, selecting a proper feature preprocessing method is also essential to further enhance the forecasting accuracy, which is the major contribution of this paper.

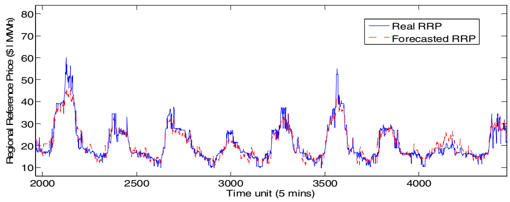

Fig. 7. Price forecast given by wrapper model + genetic search

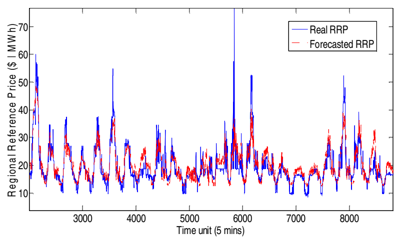

Fig. 8. Price forecast given by relief + ranker

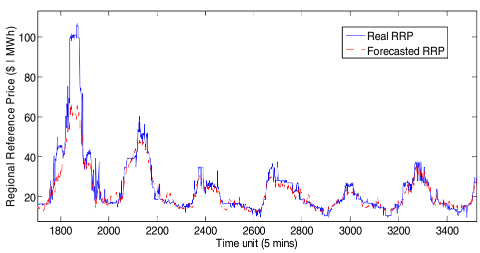

Fig. 9. Price forecast given by PCA

6 Conclusion

A systematic study of feature preprocessing techniques is presented in this paper. The authors discuss the importance of feature preprocessing in time series forecasting, as well as selecting proper feature preprocessing techniques in specific forecasting problems. Comprehensive experiments are conducted to demonstrate the effectiveness of feature preprocessing techniques with real-world electricity price data. According to the case studies, proper feature preprocessing techniques can significantly enhance the performance of time series forecasting models. It is shown in the experiment that, considering the forecasting accuracy, genetic search is the best search strategy, while wrapper model is the best evaluation criterion. Combining these techniques with a forecasting algorithm such as Support Vector Machine, a more accurate time series forecast can be provided. The main contribution of this paper exists in providing a comprehensive guidance on properly selecting effective feature preprocessing techniques and integrating them with forecasting models.

References

1. T. Kolarik, G. Rudorfer: Time Series Forecasting Using Neural Networks. Proceedings of the international conference on APL (SIGAPL'94), pages 86-94, Antwerp, Belgium, 1994.
2. J. Contreras, R. Espinola, F.J. Nogales and A.J. Conejo: ARIMA Models to Predict Next-Day Electricity Prices, IEEE Transactions on Power Systems, Vol.18, No.3, pp.1040-1020, Aug 2003.
3. D.C. Steere, A. Baptista, D. McNamee, C. Pu, and J. Walpole: Research Challenges in Environmental Observation and Forecasting Systems. Proceedings of the 6th annual international conference on Mobile computing and networking, pages 292-299, Boston, Massachusetts, USA.
4. T. Niimura and H.S. Ko: A Day-ahead Electricity Price Prediction Based on a Fuzzy-neural Autoregressive Model in a Deregulated Electricity Market. Proc. the 2002 International Joint Conf. on Neural Networks (IJCNN'02), Vol.2, No. 12-17, pp.1362-1366, May 2002.
5. J.-J. Guo and P.B. Luh: Selecting Input Factors for Clusters of Gaussian Radial Basis Function Networks to Improve Market Clearing Price Prediction, IEEE Trans on Power Systems, Vol.18, No.2, pp.665-672, May 2003.
6. H.S. Guirguis, and F.A. Felder: Further Advances in Forecasting Day-Ahead Electricity Prices Using Time Series Models, KIEE International Transactions on PE, Vol 4-A No. 3 pp. 159-166, 2004.
7. K.-R Nuller, A. J. Smola, G. Ratsch, b. Scholkopf, J. kohlmorgen, V. Vapnik: Predicting time series with support vector machine, in Proc. of ICANN'97, Springer LNCS 1327,1997. pp. 999-1004. 1997
8. Butler, K.L.; Momoh, J.A**:** Detection and classification of line faults on power distribution systems using neural networks. Proc. of the 36th Midwest Symp. on Circuits and Systems, vol. 1, 16-18 Aug. 1993 pp 368 - 371
9. H. Liu, L. Yu: Toward integrating feature selection algorithms for classification and clustering, IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 4, April, 2005.
10. Nealand, J.H.; Bradley, A.B.; Lech, M: Discriminative feature extraction applied to speaker identification, 6th Int. Conf. on Signal Processing, 2002 Vol. 1, 26-30 Aug. 2002 pp: 484 - 487.
11. A.C. Tamhane, D.D. Dunlop: Statistics and Data Analysis: from Elementary to Intermediate. Upper Saddle River, NJ: Prentice Hall, c2000.
12. A.L. Blum and P. Langley: Selection of Relevant Features and Examples in Machine Learning, Artificial Intelligence, vol. 97, pp. 245-271, 1997.
13. M. Dash, K. Choi, P. Scheuermann, and H. Liu: Feature Selection for Clustering-a Filter Solution, Proc. Second Int'l Conf. Data Mining, pp. 115-122, 2002.

14. M.A. Hall: Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning, Proc. 17th Int'l Conf. Machine Learning, pp. 359-366, 2000.
15. M. Ben-Bassat: Pattern Recognition and Reduction of Dimensionality, Handbook of Statistics-II, P.R. Krishnaiah and L.N. Kanal, eds., pp. 773-791, North Holland, 1982.
16. K. Kira and L.A. Rendell: The Feature Selection Problem: Traditional Methods and a New Algorithm, Proc. 10th Nat'l Conf. Artificial Intelligence, pp. 129-134, 1992.
17. R. Kohavi and G.H. John: Wrappers for Feature Subset Selection, Artificial Intelligence, vol. 97, nos. 1-2, pp. 273-324, 1997.
18. W.L. Martinez, A.R. Martinez: Computational Statistics Handbook with MATLAB. Boca Raton: Chapman & Hall/CRC, c2002.
19. Oja, E.; Kiviluoto, K.; Malaroiu, S: Independent component analysis for financial time series. IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC..
20. B.L. Zhang and Z.Y. Dong: An Adaptive Neural-wavelet Model for Short-Term Load Forecasting, Electric Power Systems Research, Vol. 59, pp.121–129, 2001.
21. H. Zhang, L.Z. Zhang, L. Xie, and J.N. Shen: The ANN of UMCP forecast based on developed ICA, IEEE International Conference on Electric Utility Deregulation, Reconstruction, and Power Technologies (DRPT 2004) April 2004, Hongkong.
22. Petter Skantze, Andrej Gubina, and Marija Ilic: Bid-based Stochastic Model for Electricity Prices: The Impact of Fundamental Divers on Market Dynamics. Report produced by Energy Laboratory, MIT, [Online] Available: http://lfee.mit.edu/public/el00-004.pdf.
23. J. W. Han, M. Kamber: Data Mining: Concepts and Techniques, San Francisco, Calif.: Morgan Kaufmann Publishers, 2001.
24. H. Liu and H. Motoda: Feature Selection for Knowledge Discovery and Data Mining. Boston: Kluwer Academic, 1998.
25. I.H. Witten, E. Frank: Data mining: practical machine learning tools and techniques (2nd edition). San Francisco, Calif.: Morgan Kaufman, 2005.
26. J. Yang and V. Honavar: Feature Subset Selection Using a Genetic Algorithm, Feature Extraction, Construction and Selection: A Data Mining Perspective, pp. 117-136, 1998, second printing, 2001.
27. Garcia, G.N.; Ebrahimi, T.; Vesin, J.-M: Support vector EEG classification in the Fourier and time-frequency correlation domains. Neural Engineering, 2003. Conference Proceedings. First International IEEE EMBS Conference on, 20-22 March 2003 Page(s):591 – 594.
28. T. Sakai: Average gain ratio: a simple retrieval performance measure for evaluation with multiple relevance levels. Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, July 2003.
29. V. Vapnik: The Nature of Statistical Learning Theory, Springer Verlag, New York, 1995.